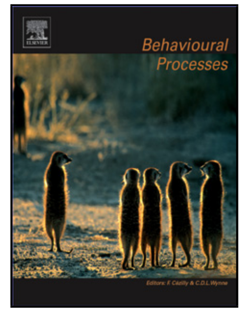
DOI:
10.1016/j.beproc.2012.11.010

# Accepted Manuscript




Title: Retrieval-induced forgetting and interference between cues: Training a cue-outcome association attenuates retrieval by alternative cues

Authors: Nerea Ortega-Castro, Miguel A. Vadillo

| | |
|---|---|
| PII: | S0376-6357(12)00245-8 |
| DOI: | doi:10.1016/j.beproc.2012.11.010 |
| Reference: | BEPROC 2572 |
| To appear in: | *Behavioural Processes* |
| Received date: | 30-7-2012 |
| Revised date: | 10-11-2012 |
| Accepted date: | 12-11-2012 |

- Retrieval-induced forgetting is a memory effect similar to interference between outcomes.
- We tried to find a similar effect with an interference-between-cues design.
- Our results show the same inhibitory effect in the two types of interference.
- RIF and both types of interference might share some common mechanisms.

**MS# BEPROC-D-12-00173R2**

Running Head: RETRIEVAL-INDUCED FORGETTING AND INTERFERENCE

**Retrieval-induced forgetting and interference between cues:**

**Training a cue-outcome association attenuates retrieval by alternative cues**

Nerea Ortega-Castro[1] & Miguel A. Vadillo[1,2]

[1]*Universidad de Deusto, Bilbao, Spain*

[2]*University College London, London, UK*

Mailing address:

Miguel A. Vadillo
Division of Psychology and Language Sciences
University College London
26 Bedford Way, London WC1H 0AH, England

Tel: +44 20 7679 5364

E-mail: m.vadillo@ucl.ac.uk

**Abstract**

Some researchers have attempted to determine whether situations in which a single cue is paired with several outcomes (A-B, A-C interference or interference between outcomes) involve the same learning and retrieval mechanisms as situations in which several cues are paired with a single outcome (A-B, C-B interference or interference between cues). Interestingly, current research on a related effect, which is known as retrieval-induced forgetting, can illuminate this debate. Most retrieval-induced forgetting experiments are based on an experimental design that closely resembles the A-B, A-C interference paradigm. In the present experiment, we found that a similar effect may be observed when items are rearranged such that the general structure of the task more closely resembles the A-B, C-B interference paradigm. This result suggests that, as claimed by other researchers in the area of contingency learning, the two types of interference, namely A-B, A-C and A-B, C-B interference, may share some basic mechanisms. Moreover, the type of inhibitory processes assumed to underlie retrieval-induced forgetting may also play a role in these phenomena.

**Keywords:** associative learning, interference, retrieval-induced forgetting

## 1. Introduction

Imagine that you are diabetic and that to live a normal life, you have no choice but to inject yourself with insulin once a day. Your doctor warns you that you should not use the same injection site each time, which can cause bruising, so he shows you a series of four points between which to alternate when injecting the drug. At first glance, this may seem to be a simple task, but the process involves both recall and forgetting. First, you must remember where the acceptable injections sites are, but at the same time your strongest memory will be of the site last used. In other words, the point on the body where you injected insulin yesterday will be most readily recalled; however, this is the same site you must avoid today. This memory is disruptive and can interfere with your goal. Therefore, you must suppress the memory and choose another injection site from the three remaining options. This example illustrates the fundamental premise of recent research regarding forgetting and remembering: to remember some important information in a given context, it is sometimes necessary to inhibit (or to not remember) related information that can be dysfunctional in that context.

Interest in forgetting has increased in recent years, resulting in a large number of new studies and debates related to its underlying mechanism (for a review, see Anderson, 2003). This has been strongly motivated by recent changes in the interpretation of forgetting, which is no longer thought of as a failure in memory, but rather as the result of an efficient and adaptive information processing mechanism (Bjork, Bjork, & Anderson, 1998). According to the most recent theories of forgetting, people selectively access necessary information needed at a given time. This control mechanism is responsible, among other things, for reducing the activation of other related memories that would be disruptive if they were recalled. Inhibiting these unwanted memories makes it easier to select the memories that are necessary in a particular situation.

The standard paradigm used to study the aforementioned inhibitory effects is illustrated in Figure 1 (see Anderson, Bjork, & Bjork, 1994). The paradigm includes a series of materials, consisting of categories and related exemplars (e.g., Fruit-orange, Fruit-banana, Vehicle-car, and Vehicle-plane), that are first presented to participants to study. In a subsequent, retrieval-practice phase, participants are asked to recall half of the exemplars of half of the categories by means of a cued-recall task (e.g., Fruit-or____). This task leads to the creation of three types of items: the RP+ items are the practiced exemplars from practiced categories (e.g., Fruit-orange), the RP- items are unpractised exemplars belonging to practiced categories (e.g., Fruit-banana), and the NRP items constitute the control condition and consist of unpractised exemplars from unpractised categories (e.g., Vehicle-car and Vehicle-plane). The third and final phase is a test phase, during which the participant must identify all of the items belonging to all of the categories that were presented in the first phase. As expected, the typical results yielded by these experiments suggest that the RP+ exemplars tend to be better remembered than the NRP items. Most interestingly however, the RP- items tend to be recalled less frequently than the NRP items, suggesting that training the RP+ exemplars results in forgetting the RP- exemplars.

Retrieval-induced forgetting (RIF) has been replicated many times using different stimuli, such as complete sentences (Anderson & Bell, 2001; Gómez-Ariza, Lechuga, Pelegrina & Bajo, 2005), personality traits (MacLeod & Macrae, 2001), crime-scene pictures (Saunders & MacLeod, 2002), and geometric shapes and colours (Ciranni & Shimamura, 1999). Additionally, RIF appears to be independent of the retrieval cues that are used during the test phase of an experiment. Many experiments test participants' memory by presenting the name of one of the trained categories and asking participants to remember the exemplars that were presented under that category name. However, other studies have reported that RIF is also observed when those exemplar names are retrieved by means of cues that were not presented during training (Anderson & Spellman, 1995). The RIF effect can even be demonstrated when participants are simply asked to indicate whether they recognise the exemplar names among a list of non-trained words (lures) or to discriminate exemplar names from non-words (Hicks & Starns, 2004). These results suggest that RIF is a general and robust effect that can be demonstrated using very different materials and testing procedures.

This phenomenon is usually interpreted in terms of a cognitive control system. From this point of view, forgetting certain information (momentarily) can be advantageous because it reduces the likelihood of dysfunctionally easy access to information in that specific context, which may facilitate the learning of new information. Most studies examining RIF assume that the inhibitory effect is driven by competition between the representation of the practiced items and the representation of the related, but unpractised, items. In classical RIF experiments, when participants are presented with the cue "Fruit", the representation of the several exemplars associated with that cue becomes active. However, only one of these exemplars is required; thus, the representations of the other exemplars must be inhibited so that they do not interfere with the retrieval of the correct item. From this point of view, the more competition there is, the more inhibition that is necessary, and it will therefore be harder to subsequently access the inhibited information. This prediction is consistent with observations that forgetting occurs more easily when the items that are potentially disruptive (i.e., the RP- exemplars) are more accessible and intrusive (e.g., words with a higher frequency; Anderson et al., 1994; see also Bäuml, Kissler, & Rak, 2002). Additionally, RIF occurs when participants are asked to actively retrieve exemplars during the retrieval practice phase, but not when they are simply presented with these exemplars for further study (Anderson & Bell, 2001). Presumably, this occurs because only in the former case must participants handle the competitive activation of conflicting representations, which triggers inhibitory processes.

As noted by Anderson (2003), the standard paradigm for studying RIF is remarkably similar to experimental designs that are traditionally used to study A-B, A-C interference or interference between outcomes (for reviews, see Postman & Underwood, 1973; Slamecka & Ceraso, 1960). The classic studies used associated pairs of words, which were studied repeatedly (A-B). Then, in a second phase, the same cue stimuli were paired with different associated responses (A-C). As a consequence, the retrievability of the association between A-B was impaired. What is interesting about this phenomenon is that the A-B, A-C paradigm is very similar to the standard RIF task. Specifically, an item is trained, such as the exemplar “orange” from the category “Fruit” (which, in this terminology, may be called A-C), which prevents subsequent recovery of another item (e.g., “banana”) that was learned together with the same category cue (which, in these terms, could be labelled as A-B).

The parallel between the experimental paradigm used to study RIF and the paradigm used to study A-B, A-C interference raises an interesting question: could a RIF-like effect be observed in a situation that is similar to traditional RIF studies with the exception that items are arranged in such a way that the task more closely resembles a different interference paradigm? Interestingly the classic studies mentioned above found that the A-B, A-C paradigm was not the only one to yield interference effects. For instance, it was also observed that pairing several cue words with a single outcome word also made it more difficult to remember the cue-outcome pairs. In other words interference is also observed in situations in which several cues are paired to a single outcome as opposed to using a single cue that is associated with several outcomes. The difference between this interference paradigm, known as A-B, C-B interference or interference between cues, and the A-B, A-C interference paradigm is represented in Figure 2. More recently, both instances of interference have drawn the attention of associative learning researchers. On one hand, many well-known associative learning effects, such as extinction and counterconditioning, refer to situations in which a given cue is paired with different outcomes at different times. These situations can therefore be described as variants of A-B, A-C interference. These effects have been extensively studied, and several theoretical interpretations have been advanced (for a review, see Bouton, 2010). On the other hand, however, some experiments have explored A-B, C-B interference effects (Matute & Pineño, 1998; see also Luque, Morís, Orgaz, Cobos, & Matute, 2011; Vadillo, Castro, Matute, & Wasserman, 2008; Vadillo, Orgaz, & Matute, 2008).

An important debate in current research is whether both types of interference are based on similar mechanisms. Interestingly, they can be observed using exactly the same materials and procedures (e.g., Escobar, Arcediano, & Miller, 2001; Pineño & Matute, 2000) and they are influenced by some manipulations in the same way. For instance, interpolating a retention interval between training and test or changing the experimental context before testing is known to reduce and even eliminate both types of interference (Pineño, Ortega, & Matute, 2000). In light of these similarities, some authors have suggested that both types of interference might share common mechanisms (Matute & Pineño, 1998, Pineño et al., 2000). However, it is remarkable that the available theories that have been proposed to explain A-B, A-C interference (e.g., Bouton, 1993, 1997) fail to account for A-B, C-B interference (but see Miller & Escobar, 2002, for a notable exception).

One unexplored possibility is that the type of inhibitory mechanisms invoked to explain RIF may play a role in both interference effects, thus providing an integrative framework. Given that the standard procedure used in RIF studies is similar to the traditional A-B, A-C interference experiments, it has been argued that both effects might be based on similar processes (Anderson, 2003). To the same extent, proving that RIF can take place in a situation similar to an A-B, C-B paradigm would provide support to the idea that A-B, C-D also involves the kind of inhibitory processes invoked by RIF researchers. Furthermore, an experiment showing that RIF can take place both when several outcomes are paired with a single cue and when several cues are paired with a single outcome would lend stronger support to the idea that both types of interference paradigms involve common mechanisms. And, most importantly, it would also suggest that the concepts and ideas developed in the study of RIF can be useful in this debate.

The main goal of the present experiment was to explore, for the first time, whether RIF can take place in a situation in which several cues are paired with a single outcome (A-B, C-B interference). Moreover, we also included an experimental condition in which one cue is paired with several outcomes (A-B, A-C interference) so that the relative strength of the RIF effects under both conditions can be compared. In this experiment, we did not use categories and exemplars as is usual in the study of RIF; instead, we used pairs of arbitrary words, namely episodic pairs, that had no semantic or causal connection whatsoever. This method allowed us to use words that could work interchangeably as either cues or outcomes without changing the nature of the semantic relationship between cues and outcomes. To the best of our knowledge, no RIF experiments have been conducted to date that use unrelated words as cues and outcomes. Therefore, a secondary goal of the present experiment was to show that significant RIF effects can be observed using these stimuli as opposed to categories and exemplars.

## 2. Method

### *2.1. Participants and Apparatus*

Ninety-four undergraduate students from the University of Deusto volunteered to take part in the experiment. None had ever participated in any related experiment. The experiment was performed in a large computer room, with each participant seated at least 1.5 m apart from adjacent participants. The experimental task was programmed using Visual Basic.

### *2.2. Stimuli*

The specific word pairs used in the experiment are presented in Appendix 1. All of the words had the same syllabic structure (beginning with a consonant) and were 6-7 letters long. Based on evidence that suggests that inhibition tends to be stronger when the stimuli to be inhibited are more intrusive (Anderson et al., 2004; Williams & Zacks, 2001), we used high-frequency words as RP- items and low-frequency words as RP+ items. The same manipulation was applied to NRP items; half of these items (NRP-) were high-frequency words, and the other half (NRP+) were low-frequency words. Therefore, half of the outcome words used in the A-B, A-C condition were high-frequency words, and the other half of the outcomes were low-frequency words. In the A-B, C-B condition, half of the cues were high-frequency words, and the other half were low-frequency words. We selected words using LEXESP (Sebastián, Cuetos, Martí, & Carreiras, 2000), using the criteria that high-frequency words ranged from 40 to 55 (mean frequency = 45) and low-frequency words ranged from 20 to 30 (mean frequency = 24.05).

Apart from the words that were used as RP+, RP-, NRP+, NRP-, and filler words that were trained and intermixed with stimuli (see below), the recognition test we used during the last phase required that the target words were presented intermixed with untrained (but comparable) words that served as lures. Twenty high-frequency words and 20 low-frequency words, which were equivalent to the experimental words in all criteria, were used as lures for the final recognition test.

As we used stimuli that had no semantic relation to each other, the items that were used as cues and outcomes were counterbalanced. This method ensured that the items assigned to the item groups (RP+, RP-, NRP+, and NRP-) were equivalent in all experimental conditions. Therefore, the items that served as outcomes in the A-B, A-C condition for some participants served as cues in the A-B, C-B condition for other participants. Similarly, the assignment of word pairs to the role of RP+ or RP- items or to the role of NRP+ or NRP- items was also counterbalanced across participants. In sum, there were four counterbalancing conditions depending on (a) whether each set of words was assigned to the A-B, C-B condition or to the A-B, C-B condition and (b) whether each subset of words was assigned to the RP+/RP- conditions or to the NRP+/NRP- conditions.

### *2.3. Procedure and Design*

#### *2.3.1. Learning phase*

The design of the experiment is summarised in Figure 3. The instructions of the experiment were shown to participants on a computer screen and explained that participants would see a series of word pairs and that their task was to learn which outcome word was paired with which cue word. After the participants reviewed these instructions, word pairs appeared on the screen, each for 4 sec with a 200-msec interval between trials. Participants were exposed to three blocks of trials, each of which contained all of the word pairs in pseudorandom order (i.e., the same randomly generated sequence of trials was used for all participants). As is usually done in most experiments on RIF, each block began and ended with three filler pairs to avoid primacy or recency effects on later recall trials of the target items.

As shown in Figure 3, two sets of eight pairs each served as the experimental condition (one for the A-B, A-C condition and the other for the A-B, C-B condition), and two similar sets of eight pairs served as the control condition (one for the A-B, A-C condition and the other for the A-B, C-B condition). Two smaller sets of four pairs served as filler items (see Figure 3 and Appendix 1). Within the experimental and control sets belonging to the A-B, A-C condition, word pairs were structured in a way such that one word was always presented as the cue for all of the pairs belonging to that condition and eight different words were presented as outcomes. In the A-B, C-B condition, the structure of the word pairs was reversed; specifically, a single word acted as the common outcome of all of the word pairs belonging to one set, and eight different words were used to cue the outcome word. One of the filler sets was structured as the A-B, A-C condition (i.e., a single cue was paired with different outcomes in different trials), and the other filler set was structured as the A-B, C-B condition (i.e., a single outcome was paired with different cues).

*2.3.2. Retrieval practice*

After the learning phase, participants were asked to retrieve half of the word pairs from half of the sets (i.e., the RP+ items). In each trial, the cue was presented for 2 sec. Then, the first two letters of the outcome word to be recalled were presented, followed by a blank box in which the participant was asked to type in the right word with no time limitation. Extant evidence suggests that this pre-cuing procedure (i.e., presenting only the cue shortly before presenting the first two letters of the outcome to be remembered) enhances the RIF effect (Bajo, Gómez-Ariza, Fernández, & Marful, 2006). Words were arranged in three blocks, each containing all of the RP+ items in a random order. At the beginning and end of each block, filler word pairs were used. None of the word pairs that were RP-, NRP-, or NRP+ items were presented in this phase. After this phase, participants were informed that they could browse the internet for 5 minutes. When 5 minutes had elapsed, participants were given a three-digit numerical code that allowed them to proceed to the last phase of the experiment.

*2.3.3. Recognition test*

Participants were exposed to each individual target word that they had studied during the first phase of the experiment in addition to an equal number of similar words that were used as lures. The target words consisted of all of the outcomes belonging to the A-B, A-C condition and all of the cues belonging to the A-B, C-B condition. Participants were asked to respond to each test trial by indicating whether or not they had seen that word in the first phase of the experiment by pressing the computer keys <Q> and <P>. If the participant responded correctly, the word on the screen changed to green, and if the answer was incorrect, the word changed to red.

## 3. Results

Visual inspection of the data revealed that some participants had an unusually high false-alarm rate in the recognition test. Some of these participants also showed very poor performance in the retrieval practice phase. Therefore, in the subsequent data analyses we removed all of the participants whose false-alarm rate was two standard deviations above the group mean or whose number of correct responses during the last block of the retrieval practice phase was two standard deviations below the group mean. Data from seven participants who met one or both of these criteria were removed. The recognition accuracy was measured with *d'* scores, computed as the difference between the (*z*-transformed) hit rate and the (*z*-transformed) false alarm rate. Hit and false-alarm rates equal to 0 or 1 were recoded as .9 or .1, respectively, to compute the *z*-transformed values necessary to calculate the *d'* recognition scores.

Table 1 presents the *d'* scores for each condition. First, we explored whether the retrieval practice phase resulted in a facilitation effect for practiced items (RP+) relative to control items. For this analysis, we contrasted recognition of the RP+ items with recognition of the NRP items that had the same (low) frequency (NRP+). A 2 (Condition: RP+ vs. NRP+) x 2 (Type of interference: A-B, A-C vs. A-B, C-B) repeated-measures analysis of variance (ANOVA) conducted on the obtained *d'* scores revealed a main effect of Condition, $F(1, 86) = 60.22$, $p < .001$, $\eta_p^2 = 0.41$, and a Condition x Type interaction, $F(1, 86) = 12.79$, $p < .001$, $\eta_p^2 = 0.13$. The main effect of Type was not significant, $F < 1$, $\eta_p^2 < 0.00$. As presented in Table 1, these results indicate that facilitation occurred in both conditions, although the effect was somewhat greater in the A-B, C-B condition. Although unexpected, this interaction is likely to be due to overall baseline differences in performance in both conditions. As can be seen in Table 1, recognition scores tended to be lower in the A-B, C-B condition. This leaves more space for facilitation to be observed in that condition, compared to condition A-B, A-C, where baseline recognition is higher, and therefore any further improvement in memory is more difficult to detect.

To assess whether RIF had also occurred, we conducted a similar analysis in which the high-frequency NPR words (NRP-) were used as a control condition for the recognition of RP- items. A 2 (Condition: RP- vs. NRP-) x 2 (Type of interference: A-B, A-C vs. A-B, C-B) repeated-measures ANOVA conducted on the *d'* scores revealed a main effect of Condition, $F(1, 86) = 4.17$, $p < .05$, $\eta_p^2 = 0.05$, which confirms that retrieval-induced forgetting occurred in both conditions. Most importantly, the Condition x Type of interference interaction was far from statistical significance, $F < 1$, $\eta_p^2 = 0.003$, which confirms that the RIF effect was not larger when several outcomes were paired with a single cue than when several cues were paired with a single outcome. We also found a main effect of Type of interference, $F(1, 86) = 17.85$, $p < .001$, $\eta_p^2 = 0.17$, which confirms that items tended to be easier to recognise in the A-B, A-C condition than in the A-B, C-B condition.

## 4. Discussion

The results of the present experiment reported here shed new light on our knowledge of RIF and interference. First, the results show that RIF can occur when stimuli such as cues and outcomes that are completely independent of each other are used, even if there is no semantic or causal relation between them. Although seemingly trivial, this is an interesting finding in its own right. As discussed above, some researchers have emphasised the relationship between RIF and classical studies on interference. However, the studies that have been conducted in both traditions utilise different materials and procedures. While most of the classical studies on interference use pairs of unrelated words, nonsense syllables or trigrams as stimuli (Osgood, 1949; Slamecka & Ceraso, 1960), to the best of our knowledge, no previous attempts have been made to induce RIF in a task based on these stimuli. Therefore, our results not only add to the body of evidence supporting the generality of RIF and its independence of the particular features of the stimuli used, but they also confirm the notion that the procedures used in interference studies are also valid for eliciting RIF. This is a very valuable piece of evidence to explain interference and RIF within an integrative framework. Moreover, as discussed below, the fact that we elicited RIF with causally unrelated stimuli also allows us to discard an important reasoning-based account of interference effects in causal learning.

From a theoretical point of view, the present experiment contributes to the debate regarding whether A-B, A-C interference and A-B, C-B interference are based on common mechanisms. In associative learning theory, interference between outcomes is usually accounted for in terms of ambiguity reduction (for a recent review of these theories, see Bouton, 2010). For instance, in a simple acquisition-and-extinction experiment, participants must learn that a cue is followed by an outcome and then, in a second phase, that the same cue is no longer followed by that outcome. This procedure makes the cue ambiguous; it can predict either the presence or the absence of the outcome. According to the most popular theories (Bouton, 1993, 1997; Gawronski, Rydell, Vervliet, & De Houwer, 2010), to solve this ambiguity, people (and other animals) use the context in which each association (cue-outcome and cue-no outcome) was learned to disambiguate the meaning of the cue. In particular, participants tend to express the association that was learned first in any context, but the conflicting association is only expressed in the particular context in which it was learned. In other words, the cue-outcome association can be generalised to novel contexts, but the cue-no outcome association is context-specific. This result means that interference will be observed whenever a cue-outcome association is retrieved in a context in which an incompatible association (i.e., cue-no outcome) was later trained. This theory provides an adequate and simple explanation of how organisms organise conflicting information and when interference can be expected. Moreover, it applies not only to extinction designs but also to a wide variety of paradigms in which a single cue is paired with several, mutually exclusive outcomes, such as latent inhibition (Lubow & Moore, 1959) and counterconditioning (Pavlov, 1927).

However, an important limitation of this theoretical framework is that it does not make any predictions about interference effects that resemble A-B, C-B interference. In these situations, several cues are paired with a common outcome (see Figure 2). Therefore, when participants are presented with one of these cues and asked to remember which outcome was paired with that cue, there is no ambiguity regarding the correct response. The "meaning" of cues is unambiguous throughout the entire experiment, and participants do not need to encode any additional information (e.g., features of the context) to better predict what outcome predicts each cue. Therefore, the standard associative learning theory that we outlined in the previous paragraph is silent about these situations.

Interestingly, there have been a number of recent attempts to account for both A-B, A-C interference and A-B, C-B interference using a common framework. For instance, Matute and Pineño (1998) and Miller and Escobar (2002) suggested that whenever two associations share a common element (either the cue or the outcome), their retrievability becomes inversely related, such that whatever primes one of the words of the pair also reduces the retrievability of the other. In the context of our experimental design, once two associations with a common outcome have been learned (e.g., Cue4-Outcome21 and Cue8-Outcome21 in Figure 3), subsequent practice with one association (e.g., Cue4-Outcome21) reduces the retrievability of the other association (Cue8-Outcome21). The same explanation extends to associations that share a common cue (e.g., Cue1-Outcome1 and Cue1-Outcome5). This account emphasises the relative retrievability of whole associations. However, our testing procedure, which included a recognition test, did not probe the retrievability of associations but rather examined the retrievability of isolated representations. In other words, our results suggest that at least part of the interference effect is due to the inhibition of specific representations of the cues or outcomes involved and not to the inhibition or the reduced retrievability of whole cue-outcome associations. Therefore, it is at best unclear whether this theoretical framework effectively explains our data.

A second framework that has been proposed to account for both types of interference is based on causal reasoning processes. In a recent series of experiments, Cobos, López, and Luque (2007; see also Luque, Cobos, & López, 2008) observed that many demonstrations of A-B, A-C and A-B, C-B interference are based on experimental paradigms in which participants can perceive a causal relationship between cues and outcomes (e.g., the cue may be a medicine, and the outcome may be an allergic reaction to that medicine). However, the causal scenario is usually different between experimental tasks that elicit A-B, A-C interference and those that elicit A-B, C-B interference. Investigations of the former tend to use experimental tasks in which cues are potential causes of the outcome.  On the other hand, investigations of the latter often utilise experimental tasks in which cues are potential effects of the outcome (e.g., the cue is a coloured light that signals whether there are hidden mines in a field, playing the role of the outcome). In other words, A-B, A-C interference is observed in predictive causal tasks, and A-B, C-B is observed in diagnostic causal tasks. Interestingly, both a predictive A-B, A-C task and a diagnostic A-B, C-B task reflect the same underlying causal structure. In both cases, a single cause is followed by different effects at different times. Therefore, this causal reasoning framework also provides an integrative explanation for both types of interference. However inspiring, we doubt that this theory can account for the results of the present experiment. Because we used unrelated words as cues and outcomes, it seems highly unlikely that participants framed the task in a causal scenario. This interpretation is consistent with recent experimental evidence showing that A-B, C-B interference can also be observed in situations in which stimuli are causally unrelated (Luque, Morís, Cobos, & López, 2009).

Given that the standard associative theory of interference between outcomes is not able to account for interference between cues and that more recent theories also seem poorly suited to account for our results, it is tempting to consider the mechanism assumed to underlie RIF as a potential explanation for both interference between outcomes and interference between cues. According to various authors, RIF is the result of a selective retrieval process that not only activates the representation of relevant information but also inhibits whatever irrelevant information comes to mind (Anderson, 2003). From this point of view, if a cue predicts several outcomes but only one of them is relevant, the representation of the potentially distracting outcomes will be inhibited. There is no reason why this mechanism cannot be extrapolated to an interference-between-cues paradigm. When several cues are paired with an outcome, presenting one cue may result in the accidental activation of the other cues that were also paired with the same outcome.[1] As these irrelevant cues may compete with the target cue and outcome for attention, their representation may also be inhibited.

The idea that the inhibitory processes involved in RIF might also play a role in interference effects receives further support from a recent experiment by Vadillo et al. (in press). In that experiment, participants were exposed to a standard preparation for the study of interference between outcomes. One cue, A, predicted an outcome, O1, in phase 1 and a different outcome, O3, during phase 2. After a brief retention interval, the participants had to learn an association between a novel cue, E, and O1. The results show that participants needed more time to learn this association between E and O1, than to learn another association between a different cue and an outcome that had not been involved in an interference treatment. These results are consistent with the idea that the representation of O1 was inhibited as a result of the interference treatment, and that this inhibition made it more difficult to learn a new association involving O1. Taken together, the results of Vadillo et al. (in press) and the results of the present experiment suggest that the inhibition of representations invoked to account RIF also plays a role in interference between outcomes and is likely to be involved in interference between cues too.

It is interesting, however, to note a crucial difference between the results observed in standard RIF studies and the results observed in recent demonstrations of interference in associative learning preparations. As mentioned in the Introduction, it is well known that interference effects seem to reduce or even vanish if a retention interval is interposed between training and testing (see Bouton, 2010; Pineño et al., 2000). However, most RIF studies include such a retention interval before testing to make sure that the observed effects are related to retrieval processes and not to encoding processes. To keep consistency with the standard procedures used in RIF research both this experiment and the one published by Vadillo et al. (in press) included a retention interval before testing. However, it is unclear whether the observed effects would be stronger or weaker if the retention interval had been briefer. Future research should explore and compare the temporal dynamics of the inhibitory processes involved in interference effects and RIF.

In any case, we think that the present experiment not only supports the generality of the RIF effect but also suggests that the mechanisms involved in this effect may play a role in phenomena that are explored in different research areas, such as associative learning. These mechanisms may also provide a unitary account of the effects that are commonly explained using contrasting theories. We hope that this work will help researchers fill the gap between the concepts used in associative learning and memory research.

## 5. References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*, 415-445.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130*, 544-570.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 1063-1087.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., & Gabrieli, J. D. E. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*, 232-235.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68-100.

Bajo, M. T., Gomez-Ariza, C. J., Fernandez, A., & Marful, A. (2006). Retrieval-induced forgetting in perceptually driven memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1185-1194.

Bäuml, K. H., Kissler, J., & Rak, A. (2002). Part-list cuing in amnesic patients: Evidence for a retrieval deficit. *Memory & Cognition, 30,* 862-870.

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103-137). Hillsdale, NJ: Erlbaum.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian conditioning. *Psychological Bulletin, 114,* 80-99.

Bouton, M. E. (1997). Signals for whether versus when an event will occur. In M. E. Bouton & M. S. Fanselow (Eds.), *Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles* (pp. 385-409). Washington, DC: American Psychological Association.

Bouton, M. E. (2010). The multiple forms of "context" in associative learning theory. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 233-258). New York: The Guilford Press.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1403-1414.

Cobos, P. L., López, F. J., & Luque, D. (2007). Interference between cues of the same outcome depends on the causal interpretation of the events. *Quarterly Journal of Experimental Psychology, 60,* 369-386.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology, 49B,* 60-80.

Escobar, M., Arcediano, F., & Miller, R. R. (2001). Conditions favoring retroactive interference between antecedent events (cue competition) and between subsequent events (outcome competition). *Psychonomic Bulletin & Review, 8,* 691-697.

Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General, 139,* 683–701.

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33,* 1431-1441.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11,* 125-130.

Lubow, R. E., y Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced preexposure to the conditioned stimulus. *Journal of Comparative and Physiological Psychology, 52,* 415-419.

Luque, D., Cobos, P. L., & López, F. J. (2008). Interference between cues requires a causal scenario: Favourable evidence for causal reasoning models in learning processes. Learning and Motivation, 39, 196–208.

Luque, D., Morís, J., Cobos, P. L., & López, F. J. (2009). Interference between cues of the same outcome in a non-causally framed scenario. Behavioural Processes, 81, 328-332.

Luque, D., Morís, J., Orgaz, C., Cobos, P. L., & Matute, H. (2011). Backward blocking and interference between cues are empirically equivalent in non-causally framed learning tasks. *Psychological Record, 61,* 141-152.

MacLeod, M. D., & Macrae, C. N. (2001). Gone today but here tomorrow: The transient nature of retrieval induced forgetting. *Psychological Science, 12,* 148-152.

Matute, H., & Pineño, O. (1998). Stimulus competition in the absence of compound conditioning. *Animal Learning & Behavior, 26,* 3-14.

Miller, R. R., & Escobar, M. (2002). Associative interference between cues and between outcomes presented together and presented apart: An integration. *Behavioural Processes, 57,* 163-185.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review, 56,* 132-143.

Pavlov, I. P. (1927). *Conditioned reflexes.* London: Clarendon Press.

Pineño, O., & Matute, H. (2000). Interference in human predictive learning when associations share a common element. *International Journal of Comparative Psychology, 13,* 16-33.

Pineño, O., Ortega, N., & Matute, H. (2000). The relative activation of associations modulates interference between elementally trained cues. *Learning and Motivation, 31,* 128-152.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19-40.

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied, 8,* 127-142.

Slamecka, N. J., & Ceraso, J. (1960). Retroactive inhibition of verbal learning. *Psychological Bulletin, 57,* 449-475.

Sebastián, N., Cuetos, F., Martí, M.A., & Carreiras, M. F. (2000). *LEXESP: Léxico informatizado del español*. Barcelona: Edicions de la Universitat de Barcelona.

Vadillo, M. A., Castro, L., Matute, H., & Wasserman, E. A. (2008). Backward blocking: The role of within-compound associations and interference between cues trained apart. *Quarterly Journal of Experimental Psychology, 61*, 185-193.

Vadillo, M. A., Orgaz, C., Luque, D., Cobos, P. L., López, F. J., & Matute, H. (in press). The role of outcome inhibition in interference between outcomes: A contingency-learning analogue of retrieval-induced forgetting. *British Journal of Psychology.*

Vadillo, M. A., Orgaz, C., & Matute, H. (2008). Overshadowed cues have reduced ability to retroactively interfere with other cues. *Learning and Motivation, 39,* 313-322.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25,* 127-151.

Williams, C. C., & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *American Journal of Psychology, 114,* 329-354

**Author Note**

Support for this research was provided by Grant IT363-10 from the Basque Government and Grant PSI2011-26965 from the Spanish Ministry of Economy and Competitiveness. Nerea Ortega-Castro was supported by fellowship BFI09.102 from the Basque Government. We would like to thank Itxaso Barberia, Pedro Cobos, Tanya Jonker, Francisco López, David Luque, Helena Matute, and Colin McLeod for their helpful comments on an earlier version of this manuscript. Correspondence concerning this article should be addressed to Miguel A. Vadillo, Division of Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AH, England. E-mail: m.vadillo@ucl.ac.uk

**Footnotes**

[1] Interestingly, this assumption would also allow some associative learning models to account for A-B, C-B interference. For instance, both Van Hamme and Wasserman (1994) and Dickinson and Burke (1996) assume that absent, but active cues modify their associate strengths. According to these models, if the representation of cue A is activated during a C-B pairing, then the associative strength of the A-B association would also be modified. Specifically, if the C-B association is strengthened and cue A is active, these models predict that the A-B association should be weakened. As discussed by Vadillo, Castro et al. (2008) this could explain several instances of A-B, C-B interference.

**Table 1**

*d' scores for each experimental condition*

| Type of Item | Condition AB-AC<br>Mean (SEM) | Condition AB-CB<br>Mean (SEM) |
|---|---|---|
| RP+ | 1.89 (.79) | 2.06 (.64) |
| RP- | 1.40 (.83) | 1.11 (.88) |
| NRP+ | 1.64 (.79) | 1.48 (.81) |
| NRP- | 1.51 (.80) | 1.28 (.82) |

*Note.* RP+ words were low-frequency words that were learned during the first phase and practiced in the second phase of the experiment. RP- words were high-frequency words that were paired with the same words as RP+ during the first phase but that were no longer practised during the second phase of the experiment. NRP+ and NRP- words correspond to low- and high-frequency words, respectively, that were used as control items for the RP+ and RP- items, respectively. They were learned during the first phase, but they were not practised during the second phase, nor were they related to any other stimuli presented during the second phase.

**Figure Captions**

*Figure 1.* Design summary of a typical retrieval-induced forgetting study.

*Figure 2.* Visual representation of the A-B, A-C and A-B, C-B interference paradigms.

*Figure 3.* Design summary of the current experiment. The cues and outcomes are semantically unrelated words.

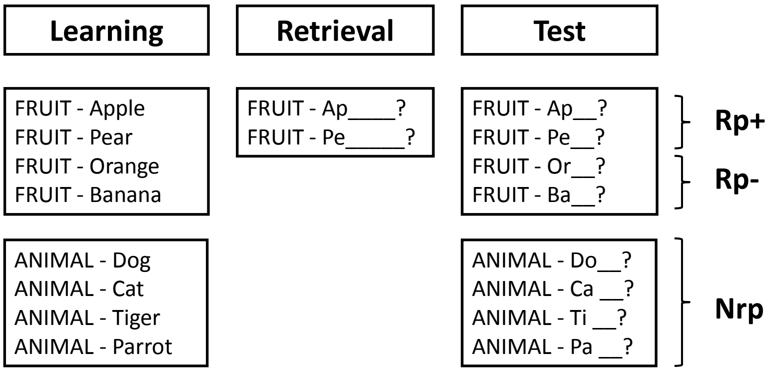

Learning
Retrieval
Test
FRUIT - Apple
FRUIT - Pear
FRUIT - Orange
FRUIT - Banana
FRUIT - Ap____?
FRUIT - Pe_____?
FRUIT - Ap__?
FRUIT - Pe__?
FRUIT - Or__?
FRUIT - Ba__?
Rp+
Rp-
ANIMAL - Dog
ANIMAL - Cat
ANIMAL - Tiger
ANIMAL - Parrot
ANIMAL - Do__?
ANIMAL - Ca __?
ANIMAL - Ti __?
ANIMAL - Pa __?
Nrp

Figure #1

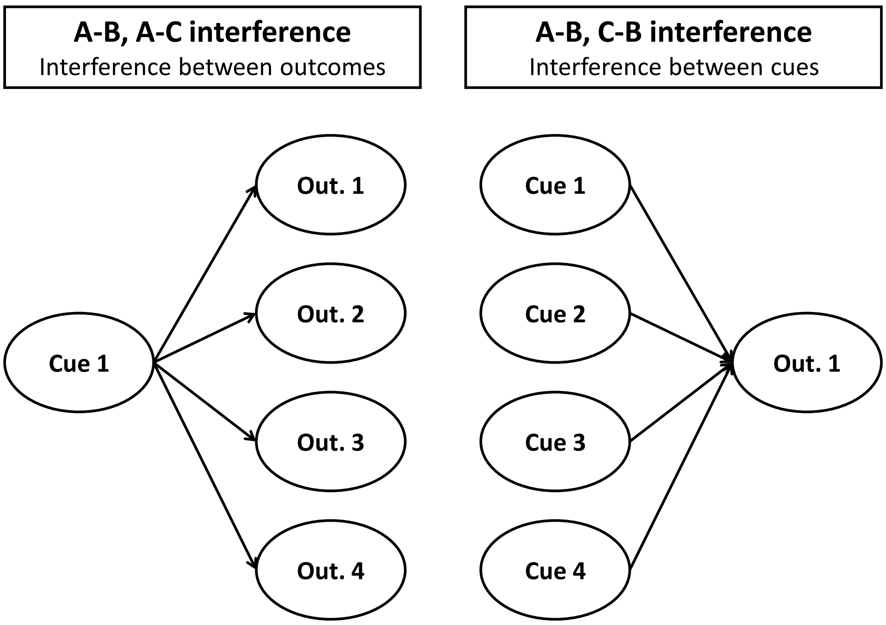

A-B, A-C interference
Interference between outcomes
A-B, C-B interference
Interference between cues
Cue 1
Out. 1
Out. 2
Out. 3
Out. 4
Cue 1
Cue 2
Cue 3
Cue 4
Out. 1

Figure #2

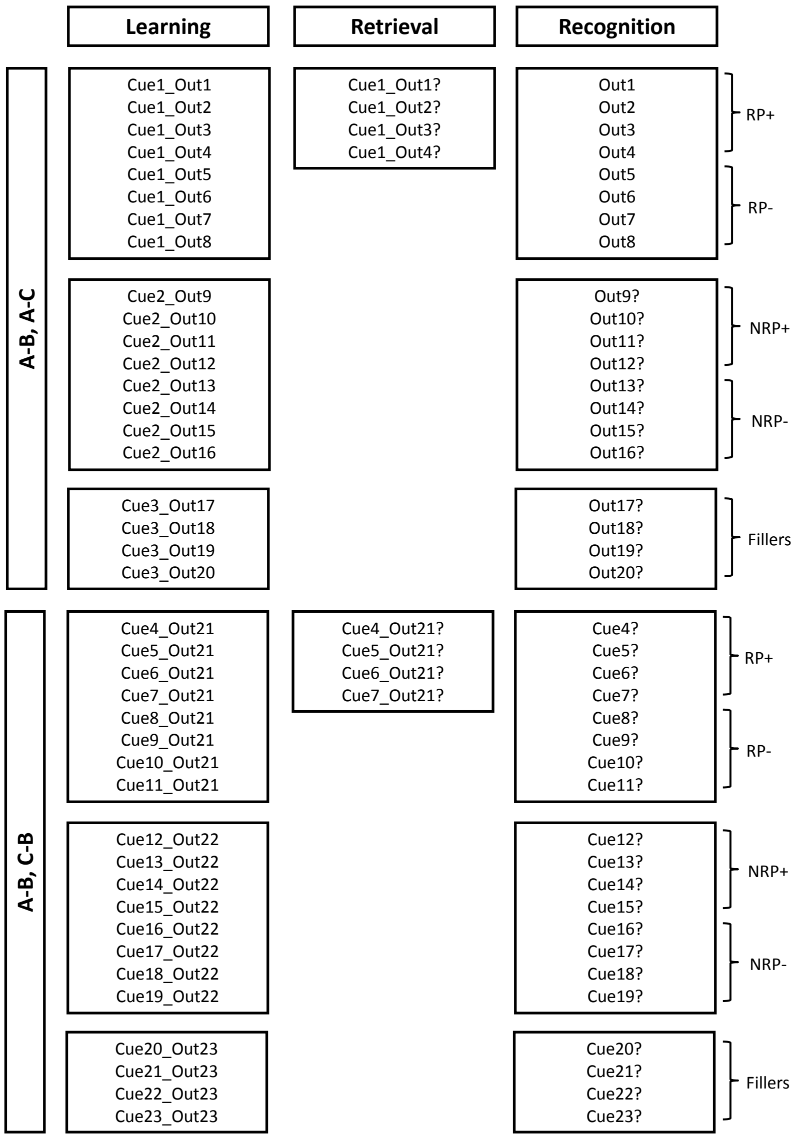

Learning
Retrieval
Recognition
A-B, A-C
Cue1_Out1
Cue1_Out2
Cue1_Out3
Cue1_Out4
Cue1_Out5
Cue1_Out6
Cue1_Out7
Cue1_Out8
Cue1_Out1?
Cue1_Out2?
Cue1_Out3?
Cue1_Out4?
Out1
Out2
Out3
Out4
Out5
Out6
Out7
Out8
RP+
RP-
Cue2_Out9
Cue2_Out10
Cue2_Out11
Cue2_Out12
Cue2_Out13
Cue2_Out14
Cue2_Out15
Cue2_Out16
Out9?
Out10?
Out11?
Out12?
Out13?
Out14?
Out15?
Out16?
NRP+
NRP-
Cue3_Out17
Cue3_Out18
Cue3_Out19
Cue3_Out20
Out17?
Out18?
Out19?
Out20?
Fillers
A-B, C-B
Cue4_Out21
Cue5_Out21
Cue6_Out21
Cue7_Out21
Cue8_Out21
Cue9_Out21
Cue10_Out21
Cue11_Out21
Cue4_Out21?
Cue5_Out21?
Cue6_Out21?
Cue7_Out21?
Cue4?
Cue5?
Cue6?
Cue7?
Cue8?
Cue9?
Cue10?
Cue11?
RP+
RP-
Cue12_Out22
Cue13_Out22
Cue14_Out22
Cue15_Out22
Cue16_Out22
Cue17_Out22
Cue18_Out22
Cue19_Out22
Cue12?
Cue13?
Cue14?
Cue15?
Cue16?
Cue17?
Cue18?
Cue19?
NRP+
NRP-
Cue20_Out23
Cue21_Out23
Cue22_Out23
Cue23_Out23
Cue20?
Cue21?
Cue22?
Cue23?
Fillers

Figure #3

**Appendix 1**

**Stimuli and general design of the experiment**

| **Experimental status** | **Word pairs in condition A-B, A-C** | **Word pairs in Condition A-B, C-B** |
|---|---|---|
| RP+ | laguna-bajada | gorila-metida |
| | laguna-dibujar | jugamos-metida |
| | laguna-calado | ribera-metida |
| | laguna-farolas | macetas-metida |
| RP- | laguna-mojado | picado-metida |
| | laguna-dedicar | doradas-metida |
| | laguna-saluda | soñaba-metida |
| | laguna-lejanas | pudimos-metida |
| NRP+ | dudaba-galope | tarima-pepita |
| | dudaba-gusanos | mitades-pepita |
| | dudaba-nevada | vereda-pepita |
| | dudaba-lunares | regalar-pepita |
| NRP- | dudaba-bodega | sujeta-pepita |
| | dudaba-separar | patadas-pepita |
| | dudaba-cometa | ceniza-pepita |
| | dudaba-vigilar | ratones-pepita |
| Filler | jinete-tirano | gemelo-cubano |
| | jinete-tenedor | tomates-cubano |
| | jinete-lozano | gitano-cubano |
| | jinete-habitan | rosales-cubano |

*Note*. This is the list of word pairs used in one of the counterbalancing conditions of the experiment. The assignment of word pairs to conditions A-B, A-C or A-B, C-B was counterbalanced across participants (reversing the order of cues and outcomes). The assignment of word pairs to the experimental (RP+, RP-) or to the control (NRP+, NRP-) conditions was also counterbalanced.